REMARKS

This is responsive to the office action issued on April 13, 2006. By this Response, claims 53, 55, 57, 65 and 71 are amended, and claims 72-75 are cancelled without prejudice. No new matter is added. Claims 53-59, 65 and 71 are active for examination.

The Examiner rejected claims 53, 55, 57, 65, 71 and 72-75 under 35 U.S.C. 112, first paragraph for failing to comply with the enablement requirement. Claim 57 was rejected under 35 U.S.C. 112, second paragraph, as being indefinite. Claims 53-59, 65 and 71-75 stood rejected under 35 U.S.C. 101 as directed to non-statutory subject matter. Claims 53-59, 65 and 71 were rejected under 35 U.S.C. 102(e) as being anticipated by Liddy (U.S. Patent No. 5,963,940). The Examiner rejected claims 72-75 under 35 U.S.C. 103(a) as unpatentable over Liddy in view of Hazlehurst (U.S. Patent No. 5,974,412).

It is submitted that the Rejections are respectfully overcome in view of the amendments and/or remarks presented herein.

The Rejections of Claims 72-75 Are Moot

By this Response, claims 72-75 are cancelled without prejudice. Accordingly, the rejections of claims 72-75 are moot.

The Rejection under 35 U.S.C. § 112, Second Paragraph Is Overcome

Claims 53, 55, 57, 65, 71 and 72-75 were rejected under 35 U.S.C. 112, first paragraph for failing to comply with the enablement requirement. Specifically, the Office Action asserted that it is not clear how the "trainable semantic vectors" of the instant invention are "trainable."

8

Applicants respectfully disagree and submit that the written description provides detailed descriptions on the trainable aspect of the claimed invention in various sections and paragraphs. For instance (the following paragraph identifications refer to paragraph numbers shown in the published patent application US 2004/0199505, the published version of the instant application), Figures 2, 8, 9, and 10 and related descriptions illustrate sample calculations from a training session; paragraph [0036], fourth sentence emphasizes the trainable aspect of the invention; paragraphs [0045] through [0056] describe the steps in the automated training process; paragraph [0105] emphasizes the trainable aspect of the invention; paragraphs [0108] through [0113] describe details of a specific training example; Paragraphs [0118] through [0125] describe the actual training that was performed in one preferred embodiment; and paragraph [0126] emphasizes the trainable aspect of the invention. For instance, a set of predetermined categories and a set of training documents are arranged such that a plurality of documents are pre-assigned to each category.

In order to address the Examiner's concern, the term "trainable" is deleted from the claims, to clarify claim scope. However, such deletion is for the purpose of reducing issues, not an admission that the specification lacks enabling description. Furthermore, the claims, as amended, specifically describe the steps for constructing semantic vectors. It is respectfully submitted that the rejection under 35 U.S.C. 112, first paragraph is overcome.

The Rejection under 35 U.S.C. § 112, Second Paragraph Is Overcome

Claim 57 was rejected under 35 U.S.C. 112, second paragraph, for lacking antecedent basis. By this Response, claim 57 is amended to provide appropriate antecedent basis. It is submitted that the rejection under 35 U.S.C. 112, second paragraph is overcome.

9

The Rejection of Claims 53-59 and 65 under 35 U.S.C. § 101 Is Overcome

Claims 1-9, 12 and 13 were rejected under 35 U.S.C. § 101 as allegedly being directed to non-statutory subject matter. The rejection is overcome.

By this Response, all the independent claims are amended. It is submitted that the claims are not directed solely to mere ideas, laws of nature, or natural phenomena. Each of the claims falls squarely into one of the classes of subject matter permitted by 35 U.S.C. § 101, that is to say process or machine, respectively. Independent claims 53 and 65, for example, are tied to machine-executed steps or a data processing system (machine). Independent claim 71 recites a tangible machine-readable medium, in conformity with *In re Beauregard*, 53 F.3d 1583, 35 USPQ2d 1383 (Fed. Cir. 1995). According to the Beauregard decision, computer programs embodied in a tangible medium, such as floppy diskettes, are patentable subject matter under 35 U.S.C. §101.

Also, as amended, each of the independent claims describes a process to efficiently identify at least one data set, such as documents, from a collection of datasets according to a query containing information indicative of desired datasets; or machine-executed steps to fin; or to efficiently identifying data points in a semantic lexicon related to a dataset. The described steps construct a semantic representation (semantic vector) for each dataset, select datasets whose semantic vectors are closest to the semantic vector for the query, and generate a result including information of the selected datasets according to a result of the selecting step. The claims also specify how semantic vectors are constructed. It is respectfully submitted that generating representative semantic vectors for datasets and identifying at least one data set from a collection of datasets according to a query containing information indicative of desired datasets, based on computer processing produces allows efficient grouping of data, which improves a machine's

10

efficiency in identifying or retrieving data according to a query. Accordingly, the claims describe process that creates a "useful, concrete and tangible result" **analogous** to the transformation of discrete dollar amounts into a final share price, which the Federal Circuit found to be a sufficiently "useful, concrete and tangible result" in *State St. Bank & Trust Co. v. Signature Fin. Group, Inc.*, 149 F.3d 1368, 47 USPQ2d 1596 (Fed. Cir. 1998). Furthermore, generating representative semantic vectors based on computer processing of information to differentiate between related and non-related datasets produces a "useful, non-abstract result" **analogous** to the method of adding a data field with information on long distance providers, which the Federal Circuit found to be a "useful, non-abstract result that facilitates *differential* billing of long-distance calls," which "fall[s] comfortably within the broad scope of patentable subject matter under §101." *AT&T Corp. v. Excel Communications, Inc.*, 172 F.3d 1352, 50 USPQ2d 1447 (Fed. Cir. 1999).

Additionally, the issuance of Liddy (U.S. Patent No. 5,963,940) and Hazlehurst (U.S. Patent No. 5,974,412) patents, both related to using computer systems to search similar documents or datasets, and cited by the Examiner to be related to the technical art described in this application, confirms that the USPTO has long considered that the technologies as claimed in the instant application are directed to patentable subject matter.

For reasons outlined above, it is submitted that all of the rejections under 35 U.S.C. § 101 should be withdrawn.

The Anticipation Rejection of Claims 53-59 and 65 Is Overcome

Claims 53-59 and 65 were rejected under 35 U.S.C. 102(e) as being anticipated by Liddy (U.S. Patent No. 5,963,940). The anticipation rejection is respectfully traversed because Liddy cannot support a prima facie case of anticipation.

11

Claim 53, as amended, describes a method for a data processing system to efficiently identify at least one data set from a collection of datasets according to a query containing information indicative of desired datasets. A semantic vector for each dataset and a semantic vector for a query are constructed. A comparison is made between the semantic vector for the query and the semantic vector of each dataset. Datasets whose semantic vectors are closest to the semantic vector for the query are selected. A result including information of the selected datasets according to a result of the selecting step is generated. The semantic vector for the query or each of the datasets is constructed by the steps of: for each data point, constructing a table for storing information indicative of a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space; determining the significance of each data point with respect to the predetermined categories; constructing a semantic vector for each data point, wherein each semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and combining the semantic vector for each of the at least one data point to form the semantic vector of the query or each of the datasets.

The use of predetermined categories corresponding to dimensions in the semantic space provides numerous benefits. The semantic dimensions for trainable semantic vectors are well known and understood ahead of time. The dimensions for trainable semantic vectors, and hence the calculations based on those dimensions, are stable over time. The trainable semantic vectors based on those dimensions are immediately readable and understandable to the end user, because they contain dimension labels corresponding to the labels on the initial categories, and further because the vector weights are on an easily interpreted scale.

12

In contrast, the approach described in Liddy is a known technique and quite different from the features described in claim 1. As correctly acknowledged by the Office Action, Liddy fails to disclose constructing a table for storing information indicative of a relationship between each data point and **predetermined categories corresponding to dimensions in the semantic space**, as described in claim 1. Since Liddy does not provide any teaching regarding predetermined categories corresponding to dimensions in the semantic space, Liddy also fails to teach the steps of determining <u>the significance of each data point with respect to the predetermined categories</u>; constructing a semantic vector for each data point, wherein each semantic vector has <u>dimensions equal to the number of predetermined categories</u> and represents the relative strength of its corresponding data point <u>with respect to each of the predetermined categories</u>; and combining the semantic vector for each of the at least one data point to form the semantic vector of the query or each of the datasets, as described in claim 1. Therefore, the anticipation rejection based on Liddy is untenable and should be withdrawn.

Other documents of record do not alleviate the deficiencies of Liddy. Contrary to the assertion of the Office Action, Hazlehurst does not disclose for each data point, constructing a table for storing information indicative of a relationship between each data point and predetermined categories corresponding to dimension in the semantic space. Hazlehurst describes the use of artificial neural networks (ANNs) and clustering techniques to automatically transform an initial collection of documents into a dictionary and further into a reduced vector space. Hazlehurst starts with just a set of documents, and then derives a new vector space whose dimensions are based solely on the mathematical operations of artificial neural networks (ANNs) and clustering algorithms. **There is no way to force these initial dimensions to align with predetermined categories**. Therefore, Hazlehurst fails to specifically support <u>predetermined</u>

13

categories corresponding to dimensions in the semantic space. Furthermore, Hazlehurst describes adjusting the vector space over time based on user feedback and learning algorithms. There is no guarantee that the new vector space will converge to precisely the desired categories. Moreover, changing the vector space over time will invalidate earlier calculations and lead to system instability over time. Additionally, it is well known that ANNs are difficult for end-users to interpret. Although ANNs can be used to automatically make decisions, they typically cannot be used to explain how those decisions were made. Furthermore, the output nodes of ANNs are not automatically labeled. Although Hazlehurst provides a small example containing vectors labeled as "vehicles", "transportation", and "tools", Hazlehurst does not teach how to produce these labels. These labels appear to be added manually for illustrative purposes, and are not automatically produced by Hazlehurst's system.'

Accordingly, like Liddy, Hazlehurst also fails to disclose constructing a table for storing information indicative of a relationship between each data point and predetermined categories **corresponding to dimensions in the semantic space**, determining the significance of each data point with respect to the predetermined categories; constructing a semantic vector for each data point, wherein each semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories; and combining the semantic vector for each of the at least one data point to form the semantic vector of the query or each of the datasets, as described in claim 1. Consequently, Liddy and Hazlehurst, even combined, do not disclose every limitation of claim 1.

Claims 57, 65 and 71 include steps for generating semantic vectors, which are substantially similar to those described in claim 1. Accordingly, claims 57, 65 and 71 also are patentable for at

14

least the same reasons as for claim 1. Favorable reconsideration of claim 1 is respectfully requested.

Claims 54-56, 58 and 59 depend on claims 53 and 57, respectively, and incorporate every limitation thereof. Therefore, claims 54-56, 58 and 59 are patentable by virtue of their dependencies. Favorable reconsideration of claims 54-56, 58 and 59 is respectfully requested.

Respectfully submitted,

MCDERMOTT, WILL & EMERY

Wei-Chen Nicholas Chen
Registration No. 56,665

600 13th Street, N.W.
Washington, DC 20005-3096
(202) 756-8000 WNC:pab
Facsimile: (202) 756-8087
Date: August 14, 2006

**Please recognize our Customer No. 20277
as our correspondence address.**